Kukush, Schneeweiss, Wolf:

Comparison of three estimators in a polynomial regression with measurement errors

Sonderforschungsbereich 386, Paper 233 (2001)

Online unter: http://epub.ub.uni-muenchen.de/

Comparison of three estimators in a polynomial regression with measurement errors

Alexander Kukush
Department of Mathematics, Kiev Taras Shevchenko University
01601 Kiev, Ukraine

Hans Schneeweiss
Institute of Statistics, University of Munich
80799 Munich, Germany

Roland Wolf
Institute of Statistics, University of Munich
80799 Munich, Germany

Abstract

In a polynomial regression with measurement errors in the covariate, which is supposed to be normally distributed, one has (at least) three ways to estimate the unknown regression parameters: one can apply ordinary least squares (OLS) to the model without regard of the measurement error or one can correct for the measurement error, either by correcting the estimating equation (ALS) or by correcting the mean and variance functions of the dependent variable, which is done by conditioning on the observable, error ridden, counter part of the covariate (SLS). While OLS is biased the other two estimators are consistent. Their asymptotic covariance matrices can be compared to each other, in particular for the case of a small measurement error variance.

Key words: Adjusted least squares, measurement errors, ordinary least squares, polynomial regression, structural least squares.

1 Introduction

The polynomial structural regression model with measurement errors in the covariate and normally distributed variables is given by the equations

$$y_i = \sum_{j=0}^{k} \beta_j \xi_i^j + \epsilon_i \tag{1}$$

$$x_i = \xi_i + \delta_i \ , \tag{2}$$

where $(\xi_i, \epsilon_i, \delta_i) \sim i.i.d.\ N((\mu_\xi, 0, 0), \Sigma)$ with $\Sigma = diag(\sigma_\xi^2, \sigma_\epsilon^2,\ \sigma_\delta^2),\ i = 1, \ldots, n$.

Two consistent estimators for the parameter vector $\beta = (\beta_0, \ldots, \beta_k)'$ have been considered in Kukush and Schneeweiss (2000): the Adjusted Least Squares (ALS) and the Structural Least Squares (SLS) estimators, see Cheng and Schneeweiss (1998) or Stefanski (1989) for ALS and Thamerus (1998) or Carroll et al. (1995) for SLS. Both assume σ_δ^2 to be known. The first one does not take the distribution of ξ into account, whereas the latter one heavily rests on the assumption that ξ is normally distributed.

As in our model ξ is indeed normally distributed by assumption, one might expect SLS to be more efficient than ALS, since SLS uses the information about the distribution of ξ while ALS does not. However, as neither of these estimators are ML estimators, this presumption is not evident at the outset and therefore needs further investigation. Some simulation results in Schneeweiss and Nittner (2000) and also in Kuha and Temple (1999) seem to indicate that SLS has in fact smaller estimation error variances than ALS. The theoretical investigations of Kukush and Schneeweiss (2000) come to the conclusion that at least in certain border line cases SLS is, in a sense, more efficient than ALS. In particular they dealt with the situation where both error variances σ_ϵ^2 and σ_δ^2 were small.

Here we take an approach which is slightly different from the one in Kukush and Schneeweiss (2000), yet leads to completely different results. We consider a model with small σ_δ^2 but comparatively large σ_ϵ^2. More precisely, we let σ_δ^2 go to zero, but leave σ_ϵ^2, together with all the other parameters of the model, constant.

It turns out that in this situation, much to our surprise, ALS and SLS are equally efficient, at least up to the order of accuracy that we here consider and which was also considered in Kukush and Schneeweiss (2000): the asymptotic covariance matrices of $\hat{\beta}_{ALS}$ and $\hat{\beta}_{SLS}$ are equal up to the order of σ_δ^2. This result can be derived using Taylor series expansions of the formulas for the asymptotic covariance matrices of $\hat{\beta}_{ALS}$ and $\hat{\beta}_{SLS}$.

In a similar way, we also derive a small-σ_δ^2 approximation to the bias and the asymptotic covariance matrix of the naive Ordinary Least Squares (OLS) estimator of β. The OLS estimator is constructed by simply using the observed, error contaminated, variable x in place of the latent, error free, variable ξ and applying the method of least squares to the polynomial regression. The OLS estimator is inconsistent, but - after a suitable normalisation - the asymptotic variances of the estimators of the components of β are less than those of the ALS or SLS estimators.

These theoretical results are corroborated by a simulation study.

In Section 2, we study the bias and the asymptotic covariance matrix of $\hat{\beta}_{OLS}$ for small σ_δ^2. In Sections 3 and 4 we do the same for $\hat{\beta}_{ALS}$ and $\hat{\beta}_{SLS}$, respectively. Section 5 compares $\hat{\beta}_{OLS}$ to $\hat{\beta}_{ALS}$ or $\hat{\beta}_{SLS}$. In Section 6 we briefly reexamine the case when both error variances are small, a case which was also studied in Kukush and Schneeweiss (2000). Section 7 presents some simulation results and Section 8 has some concluding remarks. In the appendix we prove the eventual convergence of an iterative algorithm to the unique solution of the SLS estimating equations.

2 The naive estimator $\hat{\beta}_{OLS}$

2.1 Bias and covariance matrix

With $\zeta_i = (1, \xi_i, \dots, \xi_i^k)'$ the model equation (1) can be written as

$$y_i = \zeta_i' \, \beta + \epsilon_i \ . \tag{3}$$

Replacing ζ_i with the corresponding vector $z_i = (1, x_i, \dots, x_i^k)'$ of the observable variable x, the (naive) OLS estimator of β is given by

$$\hat{\beta}_{OLS} = \left(\sum_1^n z_i z_i' \right)^{-1} \sum_1^n z_i y_i \ . \tag{4}$$

The estimating error, as derived from (3) and (4), is

$$\hat{\beta}_{OLS} - \beta = -\left(\frac{1}{n}\sum_1^n z_i z_i'\right)^{-1} \left\{\frac{1}{n}\sum_1^n z_i(z_i-\zeta_i)'\beta - \frac{1}{n}\sum_1^n z_i\epsilon_i\right\}. \quad (5)$$

By the Strong Law of Large Numbers, the bias of $\hat{\beta}_{OLS}$, $b = \lim(\hat{\beta}_{OLS} - \beta)$, is found to be [1]

$$b = -\left(Ezz'\right)^{-1} Ez(z-\zeta)'\beta \ , \quad (6)$$

and $\lim_{n\to\infty} \hat{\beta}_{OLS} = \beta + b \ \ a.s.$

In order to derive the asymptotic covariance matrix of $\hat{\beta}_{OLS}$ we use (5) and the abbreviation $S = \frac{1}{n}\sum_1^n z_i z_i'$ and consider

$$\sqrt{n}\ (\hat{\beta}_{OLS} - \beta - b) = S^{-1}\frac{1}{\sqrt{n}}\sum_1^n \{z_i\epsilon_i - z_i(z_i-\zeta_i)'\beta - z_i z_i' b\}$$

$$= S^{-1}\frac{1}{\sqrt{n}}\sum_1^n \{z_i\epsilon_i - z_i(z_i-\zeta_i)'\beta - Ezz'b - (z_i z_i' - Ezz')b\}$$

$$= S^{-1}\frac{1}{\sqrt{n}}\sum_1^n \left[z_i\epsilon_i - \{z_i(z_i-\zeta_i)' - Ez(z-\zeta)'\}\beta - (z_i z_i' - Ezz')b\right] \ ,$$

where we used (6) in the last equation. By the CLT we finally get

$$\sqrt{n}\ (\hat{\beta}_{OLS} - \beta - b) \to N(0, \Sigma_{OLS}),$$

$$\begin{aligned}\Sigma_{OLS} \quad = \quad & (Ezz')^{-1}\left[E\left\{(z-\zeta)'\beta\right\}^2 zz' - Ez(z-\zeta)'\beta\beta' E(z-\zeta)z'\right. \\ & + 2\ E\left\{(z-\zeta)'\beta\right\}(z'b)\, zz' - 2\ Ez(z-\zeta)'\beta\ b'Ezz' \\ & \left. + E(z'b)^2 zz' - Ezz'bb'Ezz'\right]\ (Ezz')^{-1} \\ & + \sigma_\epsilon^2 (Ezz')^{-1}\end{aligned}$$

With the help of (6) this can be simplified to

$$\Sigma_{OLS} = (Ezz')^{-1}\ E\left\{(z-\zeta)'\beta + z'b\right\}^2\ zz'\ (Ezz')^{-1} \ + \ \sigma_\epsilon^2 (Ezz')^{-1} \quad (7)$$

[1] We do not use brackets for the expectation operator. The operator E is always understood to operate on the whole term to the right of E, terms being separated by + or −. I.i.d. Variables without the observation index i are understood to be the variables with index i=1, say.

2.2 Expansion of bias and covariance matrix

We want to find approximate expressions for b and Σ_{OLS} as σ_δ^2 tends to zero. We have the Taylor series expansion

$$z = \zeta + \frac{d\zeta}{d\xi}\,\delta + \frac{1}{2}\,\frac{d^2\zeta}{d\xi^2}\,\delta^2 + O\left(\delta^3\right) \tag{8}$$

and therefore

$$zz' = \zeta\zeta' + \delta\left(\zeta\frac{d\zeta'}{d\xi} + \frac{d\zeta}{d\xi}\zeta'\right) + \frac{\delta^2}{2}\,\frac{d^2(\zeta\zeta')}{d\xi^2} + O\left(\delta^3\right), \tag{9}$$

where we used the identity

$$\frac{d^2(\zeta\zeta')}{d\xi^2} = 2\,\frac{d\zeta}{d\xi}\frac{d\zeta'}{d\xi} + \zeta\frac{d^2\zeta'}{d\xi^2} + \frac{d^2\zeta}{d\xi^2}\zeta'\ . \tag{10}$$

It follows that

$$Ezz' = E\zeta\zeta' + \frac{1}{2}\,\sigma_\delta^2\,E\frac{d^2(\zeta\zeta')}{d\xi^2} + O\left(\sigma_\delta^4\right) \tag{11}$$

Similarly by (8),

$$z(z-\zeta)' = \delta\zeta\,\frac{d\zeta'}{d\xi} + \delta^2\left(\frac{d\zeta}{d\xi}\frac{d\zeta'}{d\xi} + \frac{1}{2}\,\zeta\,\frac{d^2\zeta'}{d\xi^2}\right) + O\left(\delta^3\right)$$

and thus

$$Ez(z-\zeta)' = \sigma_\delta^2\,E\left(\frac{d\zeta}{d\xi}\frac{d\zeta'}{d\xi} + \frac{1}{2}\,\zeta\,\frac{d^2\zeta'}{d\xi^2}\right) + O\left(\sigma_\delta^4\right) \tag{12}$$

Finally again by (8) and using (9),

$$E\left\{(z-\zeta)'\beta\right\}^2 zz' = \sigma_\delta^2\,E\left(\frac{d\zeta'}{d\xi}\beta\right)^2\zeta\zeta' + O\left(\sigma_\delta^4\right) \tag{13}$$

With these results we are now able to present expressions for b and Σ_{OLS} which are valid up to the order of σ_δ^2. From (6), (11), and (12) we get

$$b = -\sigma_\delta^2\left(E\zeta\zeta'\right)^{-1}E\left(\frac{d\zeta}{d\xi}\frac{d\zeta'}{d\xi} + \frac{1}{2}\,\zeta\,\frac{d^2\zeta'}{d\xi^2}\right)\beta + O\left(\sigma_\delta^4\right) \tag{14}$$

In the special case of a linear regression (k=1) the well-known attenuation effect for β_1 can be seen:

$$bias(\hat{\beta}_1) = -\left(\sigma_\delta^2/\sigma_\xi^2\right)\beta_1 + O\left(\sigma_\delta^4\right).$$

Actually, in this case, the bias is given by

$$bias(\hat{\beta}_1) = -\left(\sigma_\delta^2/\sigma_x^2\right)\beta_1 \ ,$$

(see, e.g., Schneeweiss and Mittag 1986, p. 140).

Turning to the covariance matrix (7), we first note that by dropping the term $z'b$ the right hand side of (7) will change only by a term of order $O\left(\sigma_\delta^4\right)$ because, due to (14), $b = O\left(\sigma_\delta^2\right)$ and due to (8), $z - \zeta$ is of the form $a\delta + O\left(\delta^2\right)$. With the help of (11) and (13) we thus get

$$\Sigma_{OLS} = \sigma_\epsilon^2\Phi^{-1} + \sigma_\delta^2\Phi^{-1}\left\{E\left(\frac{d\zeta'}{d\xi}\beta\right)^2\zeta\zeta' - \frac{1}{2}\sigma_\epsilon^2 E\frac{d^2\left(\zeta\zeta'\right)}{d\xi^2}\right\}\Phi^{-1} + O\left(\sigma_\delta^4\right) \ , \tag{15}$$

where we introduced the abbreviation

$$\Phi = E\zeta\zeta'.$$

We used the following general result, which holds true for any positive definite matrix A and arbitrary square matrix B and for $a \to 0$:

$$(A + aB)^{-1} = A^{-1} - aA^{-1}BA^{-1} + O\left(a^2\right).$$

3 The adjusted least squares estimator $\hat{\beta}_{ALS}$

3.1 Covariance matrix

The adjusted least squares estimator of β is the (a.s. unique) solution of the unbiased estimating equation

$$\sum_1^n H_i\ \hat{\beta}_{ALS} = \sum_1^n h_i \ , \tag{16}$$

where $(H_i)_{rs} = t_{r+s}(x_i),\ r,s = 0,\ldots,k,\ (h_i)_r = y_i t_r(x_i),\ r = 0,\ldots,k,$ with $t_r(x)$ being a polynomial in x of degree r such that for $x = \xi + \delta$

$$E\left\{t_r(x) \mid \xi\right\} = \xi^r. \tag{17}$$

The polynomials $t_r(x)$ satisfy the recursion formula

$$t_{r+1}(x) = x\ t_r(x) - \sigma_\delta^2\ r\ t_{r-1}(x) \tag{18}$$

with initial conditions $t_0(x) = t_{-1}(x) = 0$, c.f. Cheng and Schneeweiss (1998). σ_δ^2 is assumed to be known.

According to the theory of unbiased estimating equations the estimator $\hat{\beta}_{ALS}$ is asymptotically normally distributed with asymptotic covariance matrix

$$\Sigma_{ALS} = (EH)^{-1}\, E(H\beta - h)(H\beta - h)'\, (EH)^{-1}\,. \tag{19}$$

We obviously have $E(H|\xi) = \zeta\zeta'$ and thus

$$EH = E\zeta\zeta' = \Phi\ . \tag{20}$$

Let $t = (t_0(x), \dots, t_k(x))'$, then

$$h = yt = t\zeta'\beta + \epsilon t\ ,$$
$$H\beta - h = (H - t\zeta')\ \beta - \epsilon t\ ,$$

and thus

$$E(H\beta - h)(H\beta - h)' = E(H - t\zeta')\beta\beta'(H - \zeta t') + \sigma_\epsilon^2\ Ett'.$$

The asymptotic covariance matrix can therefore be written as

$$\Sigma_{ALS} = \Phi^{-1} \left\{\sigma_\epsilon^2\ Ett' + E(H - t\zeta')\beta\beta'(H - \zeta t')\right\} \Phi^{-1} \tag{21}$$

3.2 Expansion of Σ_{ALS}

From the recursion formula (18) it follows that

$$t = z - \frac{1}{2}\ \sigma_\delta^2\ \frac{d^2 z}{dx^2} + O\left(\sigma_\delta^4\right)\ ,$$

and by (8) and a corresponding expansion for $\frac{d^2 z}{dx^2}$ we have

$$t = \zeta + \frac{d\zeta}{d\xi}\ \delta + \frac{1}{2}\ \frac{d^2\zeta}{d\xi^2}\left(\delta^2 - \sigma_\delta^2\right) + R_1\ , \tag{22}$$

where $R_1 = \sigma_\delta^2\ O\left(\delta\right) + O\left(\delta^3\right) + O\left(\sigma_\delta^4\right)$. All the following remainder terms R_j will be of the same form.

Because the elements of H consist of the polynomials $t_r(x)$, we have an analogous expansion for H:

$$H = \zeta\zeta' + \frac{d(\zeta\zeta')}{d\xi}\ \delta + \frac{1}{2}\frac{d^2(\zeta\zeta')}{d\xi^2}\ \left(\delta^2 - \sigma_\delta^2\right) + R_2\ .$$

From (22) follows

$$t\zeta' = \zeta\zeta' + \frac{d\zeta}{d\xi}\ \zeta'\delta + \frac{1}{2}\ \frac{d^2\zeta}{d\xi^2}\ \zeta'\left(\delta^2 - \sigma_\delta^2\right) + R_3$$

and thus

$$H - t\zeta' = \zeta\ \frac{d\zeta'}{d\xi}\delta + A\left(\delta^2 - \sigma_\delta^2\right) + R_4\ ,$$

where A is a square matrix. It follows that

$$E\left(H - t\zeta'\right)\beta\beta'\left(H - \zeta t'\right) = \sigma_\delta^2\ E\left(\frac{d\zeta'}{d\xi}\beta\right)^2\zeta\zeta' + O\left(\sigma_\delta^4\right) \qquad (23)$$

From (22) it also follows that

$$Ett' = E\zeta\zeta' + \sigma_\delta^2\ E\frac{d\zeta}{d\xi}\frac{d\zeta'}{d\xi} + O\left(\sigma_\delta^4\right) \qquad (24)$$

Substituting (23) and (24) into (21), we finally get

$$\Sigma_{ALS} = \sigma_\epsilon^2\Phi^{-1} + \sigma_\delta^2\ \Phi^{-1}\left\{E\left(\frac{d\zeta'}{d\xi}\beta\right)^2\zeta\zeta' + \sigma_\epsilon^2\ E\frac{d\zeta}{d\xi}\frac{d\zeta'}{d\xi}\right\}\Phi^{-1} + O\left(\sigma_\delta^4\right) \qquad (25)$$

4 The structural least squares estimator $\hat{\beta}_{SLS}$

4.1 Covariance matrix

We can reformulate (1) as a mean-variance model in the latent variable ξ:

$$E(y|\xi) = \zeta'\beta$$
$$V(y|\xi) = \sigma_\epsilon^2.$$

We find a new (conditional) mean-variance model in the observable variable x by taking conditional expectations given x:

$$E(y|x) = \mu(x)'\beta \overset{df}{=} m(x,\beta) \tag{26}$$

$$V(y|x) = \sigma_\epsilon^2 + \beta'(M(x) - \mu(x)\mu(x)')\beta \overset{df}{=} v(x,\beta,\sigma_\epsilon^2) \tag{27}$$

Here $\mu(x) = (1, \mu_1(x), \ldots, \mu_k(x))'$ is a vector consisting of components

$$\mu_r(x) = E(\xi^r|x)$$

and $(M(x))_{r,s} = \mu_{r+s}(x)$, $r, s = 0, \ldots, k$.
Now, because of (2), the conditional distribution of ξ given x is $N(\mu_1(x), \tau^2)$ with

$$\mu_1(x) = \mu_x + (1 - \sigma_\delta^2/\sigma_x^2)\ (x - \mu_x) \tag{28}$$

$$\tau^2 = \sigma_\delta^2\ (1 - \sigma_\delta^2/\sigma_x^2) \tag{29}$$

We can therefore compute the conditional moments of ξ given x as

$$\mu_r(x) = \sum_{j=0}^{r} \binom{r}{j}\ \mu_j^*\ \mu_1(x)^{r-j} \tag{30}$$

$$\mu_j^* = \begin{cases} 0 & \text{if j is odd} \\ (j-1)\,!\,!\ \tau^j & \text{if j is even} \end{cases} .$$

In these formulas the nuisance parameters μ_x and σ_x^2 have to be replaced with their estimates in order to derive a mean-variance model corresponding to (26), (27), in which $\hat{\beta}_{SLS}$ and $\sigma_{\epsilon SLS}^2$ are then found as the solution to the (asymptotically unbiased) estimating equations

$$\sum_1^n \frac{y_i - \hat{m}(x_i,\beta)}{\hat{v}(x_i,\beta,\sigma_\epsilon^2)}\ \hat{\mu}(x_i) = 0$$

$$\sigma_\epsilon^2 = \frac{1}{n-k-1}\sum_1^n \{y_i - \hat{m}(x_i,\beta)\}^2 - \frac{1}{n}\sum_1^n \beta' \left\{\hat{M}(x_i) - \hat{\mu}(x_i)\ \hat{\mu}(x_i)'\right\}\ \beta\ ,$$

c.f. Kukush and Schneeweiss (2000); the second equation is different from (15) in Kukush and Schneeweiss (2000), but can just as well be used to estimate and update σ_ϵ^2. The "hat" serves to remind us of the replacement of μ_x and σ_x^2 with their estimates $\bar{x}$ and s_x^2.

The estimating equations are solved by an iterative procedure (iteratively reweighted least squares). In the appendix this procedure is explained, and it is shown that it converges eventually, i.e., for sufficiently large n. During the iterations and also as a final result $\hat{\sigma}_\epsilon^2$ may become negative. The estimation method can be greatly improved by providing bounds for $\hat{\sigma}_\epsilon^2$, say $\frac{1}{n}\hat{\sigma}_{\epsilon OLS}^2 < \hat{\sigma}_\epsilon^2 < n\hat{\sigma}_{\epsilon OLS}^2$. Whenever one of these bounds is exceeded, the estimate $\hat{\sigma}_\epsilon^2$ is set equal to that bound.

The resulting estimators of β and σ_ϵ^2 are consistent and asymptotically normally distributed. In particular, the asymptotic covariance matrix Σ_{SLS} of $\hat{\beta}_{SLS}$ is given in Kukush and Schneeweiss (2000), (35). From the discussion in 5.2.2 of that paper it can be seen that the terms in Σ_{SLS} originating from the estimation of the nuisance parameters are of the order $O\left(\sigma_\delta^4\right)$ and the formula for Σ_{SLS} therefore simplifies to

$$\Sigma_{SLS} = \left(E\frac{\mu(x)\ \mu(x)'}{v(x)}\right)^{-1} + O\left(\sigma_\delta^4\right) \tag{31}$$

where we dropped the dependence on β and σ_ϵ^2 in the notation of $v(x)$.

4.2 Expansion of Σ_{SLS}

As we want to express Σ_{SLS} in terms of the parameters of the distribution of ξ rather than of x, we first note that $\mu_x = \mu_\xi$ and $\sigma_x^2 = \sigma_\xi^2 + \sigma_\delta^2$.

Expanding (29), (28) and (30), we get

$$\begin{aligned}
\tau^2 &= \sigma_\delta^2 + O\left(\sigma_\delta^4\right) \\
\mu_1(x) &= \xi + \delta - \frac{\sigma_\delta^2}{\sigma_\xi^2}(\xi - \mu_\xi) + R_5 \\
\mu_r(x) &= \mu_1(x)^r + \binom{r}{2}\tau^2\mu_1(x)^{r-2} + R_6 \\
&= \xi^r + r\ \xi^{r-1}\ \delta + \binom{r}{2}\xi^{r-2}\delta^2 - r\ \xi^{r-1}\ \frac{\sigma_\delta^2}{\sigma_\xi^2}\ (\xi - \mu_\xi) \\
&\quad + \binom{r}{2}\sigma_\delta^2\ \xi^{r-2} + R_7\ ,
\end{aligned}$$

where R_5, R_6, R_7 (and the following R_j as well) are of the same form as, e.g., R_1 in (22). The latter equation, for $r = 0, \ldots, k$, can also be written in vector form:

$$\mu = \mu(x) = \zeta + \delta\ \frac{d\zeta}{d\xi} + \frac{\delta^2}{2}\ \frac{d^2\zeta}{d\xi^2} - \sigma_\delta^2\ \frac{\xi - \mu_\xi}{\sigma_\xi^2}\ \frac{d\zeta}{d\xi} + \frac{1}{2}\ \sigma_\delta^2\ \frac{d^2\zeta}{d\xi^2} + R_8\ . \quad (32)$$

From this we get

$$\begin{aligned} \mu\mu' &= \zeta\zeta' + \delta\ \frac{d(\zeta\zeta')}{d\xi} + \frac{\delta^2}{2}\ \frac{d^2(\zeta\zeta')}{d\xi^2} \\ &\quad + \sigma_\delta^2 \left\{ \frac{1}{2} \left(\zeta \frac{d^2\zeta'}{d\xi^2} + \frac{d^2\zeta}{d\xi^2}\zeta' \right) - \frac{\xi - \mu_\xi}{\sigma_\xi^2}\ \frac{d(\zeta\zeta')}{d\xi} \right\} + R_9 \quad (33) \end{aligned}$$

where we used again the identity (10) and in addition the similar identity

$$\frac{d(\zeta\zeta')}{d\xi} = \zeta\ \frac{d\zeta'}{d\xi} + \frac{d\zeta}{d\xi}\ \zeta'.$$

Now the elements of $M(x)$ are the same $\mu_r(x)$ as the elements of $\mu(x)$. Therefore we have an expansion for $M(x)$ analogous to that of $\mu(x)$ in (32):

$$\begin{aligned} M = M(x) &= \zeta\zeta' + \delta\ \frac{d(\zeta\zeta')}{d\xi} + \frac{\delta^2}{2}\ \frac{d^2(\zeta\zeta')}{d\xi^2} - \sigma_\delta^2\ \frac{\xi - \mu_\xi}{\sigma_\xi^2}\ \frac{d(\zeta\zeta')}{d\xi} \\ &\quad + \frac{1}{2}\ \sigma_\delta^2\ \frac{d^2(\zeta\zeta')}{d\xi^2} + R_{10}. \end{aligned}$$

So finally, again using (10), we get

$$M - \mu\mu' = \sigma_\delta^2\ \frac{d\zeta}{d\xi}\frac{d\zeta'}{d\xi} + R_{11}.$$

Now substituting this result into (27), we find

$$v = v(x) = \sigma_\epsilon^2 + \sigma_\delta^2 \left(\frac{d\zeta'}{d\xi}\beta \right)^2 + R_{12} \quad (34)$$

and

$$\frac{\mu\mu'}{v} = \frac{1}{\sigma_\epsilon^2}\ \mu\mu' \left\{ 1 - \frac{\sigma_\delta^2}{\sigma_\epsilon^2} \left(\frac{d\zeta'}{d\xi}\beta \right)^2 \right\} + R_{13}$$

$$= \frac{1}{\sigma_\epsilon^2}\left\{\mu\mu' - \frac{\sigma_\delta^2}{\sigma_\epsilon^2}\left(\frac{d\zeta'}{d\xi}\beta\right)^2 \zeta\zeta'\right\} + R_{14} \ ,$$

where the last equation follows from (33). Taking expectations and using (33) again, we derive

$$E\frac{\mu\mu'}{v} = \frac{1}{\sigma_\epsilon^2}\left[\Phi + \frac{\sigma_\delta^2}{2}\left\{E\left(\frac{d^2(\zeta\zeta')}{d\xi^2} + \zeta\frac{d^2\zeta'}{d\xi^2} + \frac{d^2\zeta}{d\xi^2}\zeta'\right) - 2A\right\} - \frac{\sigma_\delta^2}{\sigma_\epsilon^2}\ E\left(\frac{d\zeta'}{d\xi}\beta\right)^2 \zeta\zeta'\right] + O\left(\sigma_\delta^4\right) \ , \quad (35)$$

where

$$A = E\frac{\xi - \mu_\xi}{\sigma_\xi^2}\ \frac{d(\zeta\zeta')}{d\xi} \ ,$$

which will now be further analyzed.

Let $\xi_* = (\xi - \mu_\xi)/\sigma_\xi$. Then

$$E\xi_*^{r+1} = r\ E\xi_*^{r-1} \ ,\ r = 1, 2, \ldots$$

because $\xi_* \sim N(0,1)$. By binomial expansion it follows that

$$E\left(\xi_* + a\right)^r \xi_* = r\ E\left(\xi_* + a\right)^{r-1} .$$

Applying this to $a = \mu_\xi/\sigma_\xi$ we get

$$E\xi^r \frac{\xi_*}{\sigma_\xi} = r\ E\xi^{r-1} = E\frac{d\xi^r}{d\xi} \ .$$

As the elements of $\frac{d(\zeta\zeta')}{d\xi}$ are of the form ξ^r, we can apply this result to the expression A and find

$$A = E\frac{d(\zeta\zeta')}{d\xi}\ \frac{\xi_*}{\sigma_\xi} = E\frac{d^2(\zeta\zeta')}{d\xi^2} \ .$$

Substituting this expression for A into (35) and again using (10) we get

$$E\frac{\mu\mu'}{v} = \frac{1}{\sigma_\epsilon^2}\left[\Phi - \sigma_\delta^2\ E\frac{d\zeta}{d\xi}\frac{d\zeta'}{d\xi} - \frac{\sigma_\delta^2}{\sigma_\epsilon^2}\ E\left(\frac{d\zeta'}{d\xi}\beta\right)^2 \zeta\zeta'\right] + O\left(\sigma_\delta^4\right) .$$

Finally according to (31), Σ_{SLS} can be approximated by

$$\Sigma_{SLS} = \sigma_\epsilon^2 \; \Phi^{-1} + \sigma_\delta^2 \; \Phi^{-1} \left\{ E \left(\frac{d\zeta'}{d\xi} \beta \right)^2 \zeta\zeta' + \sigma_\epsilon^2 \; E \frac{d\zeta}{d\xi} \frac{d\zeta'}{d\xi} \right\} \Phi^{-1} + O\left(\sigma_\delta^4\right),$$

which is the same expression as (25).

We thus can formulate the main result of the paper:

For $\sigma_\delta^2 \to 0$,

$$\Sigma_{ALS} = \Sigma_{SLS} + O\left(\sigma_\delta^4\right).$$

5 Comparison of Σ_{OLS} to Σ_{ALS} and Σ_{SLS}

As Σ_{ALS} and Σ_{SLS} are equal up to the order of σ_δ^2 we need only consider the difference of (25) and (15):

$$\Sigma_{ALS} - \Sigma_{OLS} = \sigma_\delta^2 \; \sigma_\epsilon^2 \; \Phi^{-1} \left\{ E \frac{d\zeta}{d\xi} \frac{d\zeta'}{d\xi} + \frac{1}{2} \; E \frac{d^2(\zeta\zeta')}{d\xi^2} \right\} \Phi^{-1} + O\left(\sigma_\delta^4\right).$$

Unfortunately it is not true in general that the matrix in braces - call it B - is positive semidefinite. Indeed for $k \geq 2$ and $\xi \sim N(0,1)$, the submatrix of B consisting of the first three rows and columns is

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 10 \end{pmatrix},$$

which is indefinite, as it has a negative determinant. Thus, e.g., for the linear combination $\alpha = 4\beta_0 + 3\beta_2$ the OLS estimator has a larger asymptotic variance than the ALS estimator up to the order of σ_δ^2. More precisely:

$$\sigma^2\left(\hat{\alpha}_{OLS}\right) - \sigma^2\left(\hat{\alpha}_{ALS}\right) = 2 \; \sigma_\delta^2 \sigma_\epsilon^2 + O(\sigma_\delta^4).$$

However, the diagonal elements of B are obviously all positive except for the first one, which is zero. This means that the difference of the asymptotic variances of the ALS and SLS estimators of the j-th element of $\Phi\beta$, $j = 1, \ldots, k$, is always positive for small enough σ_δ^2.

In this sense, OLS is superior to ALS or SLS, as far as the asymptotic variances go. It should be remembered, however, that OLS is a biased estimator.

6 Case of small σ_δ^2 and σ_ϵ^2

In this section we briefly return to a case dealt with in the earlier discussion paper Kukush and Schneeweiss (2000) where σ_δ^2 and σ_ϵ^2 both tend to zero. The results of that earlier paper can now be derived very simply. We make, however, the simplifying assumption that σ_δ^2 and σ_ϵ^2 remain proportional to each other while going to zero, i.e.:

$$\sigma_\delta^2 = \kappa^2 \sigma_\epsilon^2 \tag{36}$$

with some fixed constant $\kappa > 0$. We also derive a formula for the asymptotic covariance matrix of $\hat{\beta}_{OLS}$ not considered in Kukush and Schneeweiss (2000).

6.1 Expansion of Σ_{OLS} and Σ_{ALS}

As in the various approximation formulas that were used in the derivation of (15) only δ or σ_δ^2 was involved, (15) remains essentially unchanged except that the last term in braces multiplied by σ_δ^2 is $O\left(\sigma_\delta^4\right)$ and can therefore be dropped. We thus have

$$\Sigma_{OLS} = \sigma_\epsilon^2 \; \Phi^{-1} \left\{ \Phi + \kappa^2 \; E \left(\frac{d\zeta'}{d\xi} \beta \right)^2 \zeta\zeta' \right\} \Phi^{-1} + O\left(\sigma_\delta^4\right).$$

If we introduce the term

$$v_0 = v_0(\xi, \beta) = 1 + \kappa^2 \left(\frac{d\zeta'}{d\xi} \beta \right)^2 \tag{37}$$

we can write Σ_{OLS} as

$$\Sigma_{OLS} = \sigma_\epsilon^2 \; \Phi^{-1} \; E v_0 \zeta\zeta' \; \Phi^{-1} + O\left(\sigma_\delta^4\right) \; . \tag{38}$$

The same arguments can be applied to the derivation of (25). Again the last term in braces can be dropped and consequently Σ_{ALS} will be the same as Σ_{OLS} up to the order considered:

$$\Sigma_{ALS} = \Sigma_{OLS} + O\left(\sigma_\delta^4\right).$$

6.2 Expansion of Σ_{SLS}

In the expansion (33) for $\mu\mu'$ we need only the first term:

$$\mu\mu' = \zeta\zeta' + O(\delta) + O\left(\sigma_\delta^2\right). \tag{39}$$

The expansion (34) for v is taken over, but with σ_δ^2 replaced by $\kappa^2\sigma_\epsilon^2$, so that, using the abbreviation (37), we have:

$$v = \sigma_\epsilon^2 \left\{ 1 + \kappa^2 \left(\frac{d\zeta'}{d\xi}\beta \right)^2 \right\} + R_{12} = \sigma_\epsilon^2 \; v_0 + R_{12} \tag{40}$$

with $R_{12} = \sigma_\delta^2 \; O(\delta) + O\left(\delta^3\right) + O\left(\sigma_\delta^4\right)$.
Equations (39) and (40) imply

$$\frac{\mu\mu'}{v} = \frac{1}{\sigma_\epsilon^2} \left\{ \frac{\zeta\zeta'}{v_0} + O(\delta) + \frac{O\left(\delta^3\right)}{\sigma_\delta^2} + O\left(\sigma_\delta^2\right) \right\} ,$$

hence

$$E\frac{\mu\mu'}{v} = \frac{1}{\sigma_\epsilon^2} \left\{ E\frac{\zeta\zeta'}{v_0} + O\left(\sigma_\delta^2\right) \right\} ,$$

and finally

$$\begin{aligned} \left(E\frac{\mu\mu'}{v} \right)^{-1} &= \sigma_\epsilon^2 \left\{ \left(E\frac{\zeta\zeta'}{v_0} \right)^{-1} + O\left(\sigma_\delta^2\right) \right\} \\ &= \sigma_\epsilon^2 \left(E\frac{\zeta\zeta'}{v_0} \right)^{-1} + O\left(\sigma_\delta^4\right). \end{aligned}$$

Therefore by (31),

$$\Sigma_{SLS} = \sigma_\epsilon^2 \left(E\frac{\zeta\zeta'}{v_0} \right)^{-1} + O\left(\sigma_\delta^4\right). \tag{41}$$

Thus if σ_δ^2 and σ_ϵ^2 both go to zero Σ_{ALS} and Σ_{SLS} differ at the order $O\left(\sigma_\delta^2\right) = O\left(\sigma_\epsilon^2\right)$, and SLS turns out to be more efficient than ALS in this case, see Kukush and Schneeweiss (2000), Theorem 3.

7 Some simulation results

We simulated a quadratic model with $\beta = (0, 1, -0.5)'$ and a cubic model with $\beta = (0, 1, -0.5, 0.5)'$. The sample size was taken to be $n = 200$ and $n = 900$. We let $\xi \sim N(0, 1)$. The error variances were set equal in Schneeweiss and Nittner (2000). Here, however, we studied two different scenarios. In the first one we took $\sigma_\epsilon^2 = 20$ and $\sigma_\delta^2 = 0.05$, so that the theoretical results of Sections 2 to 5, which are approximations for small σ_δ^2 but not so small σ_ϵ^2, would come out more clearly. The other scenario deals with the case of Section 6, when both error variances are small. We took $\sigma_\epsilon^2 = 0.002$ and $\sigma_\delta^2 = 0.01$. The number of replications was $N = 1000$.

We replaced ALS by a modified ALS method, called MALS. MALS has the same asymptotic properties as ALS, but is much stabler for small n, see Cheng et al. (2000).

We computed bias and standard error of the various estimators directly from the 1000 replications and compared them to the theoretical approximation values as computed from (14), (15), and (25) in the first scenario and (14) and (38) in the second scenario. The results corroborate the theory, especially when the sample size is large ($n = 900$), but to a large extent also in the case of smaller sample size ($n = 200$). Naturally, the simulation results are less stable for the cubic regression, but even there they agree sufficiently well with the theoretical results.

In the second scenario, we also computed the differences of the empirical covariance matrices of the three estimators. They were standardized by multiplying them with n/σ_δ^2. The results agree well with the theoretical statements in that the difference $\Sigma_{ALS} - \Sigma_{OLS}$ is very small, whereas $\Sigma_{ALS} - \Sigma_{SLS}$ is large.

Quadratic model: $\beta = (0, 1, -0.5)', \mu_\xi = 0, \sigma_\xi^2 = 1, \sigma_\varepsilon^2 = 20, \sigma_\delta^2 = 0.05$

Bias

	OLS		MALS	SLS
$n = 200$	theoretic	simulation	simulation	simulation
β_0	-0.025	-0.0431	-0.0202	-0.0197
β_1	-0.050	-0.0661	-0.0213	-0.0190
β_2	0.050	0.0552	0.0095	0.0088
$n = 900$	theoretic	simulation	simulation	simulation
β_0	-0.025	-0.02237	0.0016	0.0017
β_1	-0.050	-0.05303	-0.0061	-0.0054
β_2	0.050	0.04375	-0.0033	-0.0034

Standard Deviation

	OLS		MALS	SLS	
$n = 200$	theoretic	simulation	simulation	simulation	theoretic
β_0	0.3883	0.3874	0.3943	0.3939	0.3947
β_1	0.3098	0.3052	0.3207	0.3210	0.3256
β_2	0.2139	0.2169	0.2403	0.2402	0.2361
$n = 900$	theoretic	simulation	simulation	simulation	theoretic
β_0	0.1830	0.1838	0.1871	0.1872	0.1860
β_1	0.1461	0.1477	0.1553	0.1553	0.1535
β_2	0.1008	0.1018	0.1128	0.1125	0.1113

Cubic model: $\beta = (0, 1, -0.5, 0.5)', \mu_\xi = 0, \sigma_\xi^2 = 1, \sigma_\varepsilon^2 = 20, \sigma_\delta^2 = 0.05$

Bias

	OLS		MALS	SLS
$n = 200$	theoretic	simulation	simulation	simulation
β_0	-0.025	-0.0201	0.0048	0.0034
β_1	0.025	0.0096	-0.0728	-0.0132
β_2	0.050	0.0465	-0.0021	0.0000
β_3	-0.075	-0.0635	0.0311	0.0068
$n = 900$	theoretic	simulation	simulation	simulation
β_0	-0.025	-0.0292	-0.0046	-0.0043
β_1	0.025	0.0190	-0.0192	-0.0035
β_2	0.050	0.0495	0.0023	0.0019
β_3	-0.075	-0.0685	0.0068	0.0006

Standard Deviation

	OLS		MALS	SLS	
$n = 200$	theoretic	simulation	simulation	simulation	theoretic
β_0	0.3972	0.4151	0.4256	0.4241	0.4035
β_1	0.5325	0.5352	0.6049	0.5846	0.5689
β_2	0.2421	0.2509	0.2869	0.2777	0.2619
β_3	0.1497	0.1587	0.2004	0.1858	0.1656
$n = 900$	theoretic	simulation	simulation	simulation	theoretic
β_0	0.1873	0.1876	0.1920	0.1911	0.1902
β_1	0.2510	0.2376	0.2674	0.2575	0.2682
β_2	0.1141	0.1137	0.1276	0.1242	0.1235
β_3	0.0706	0.0669	0.0821	0.0769	0.0780

Quadratic model: $\beta = (0, 1, -0.5)', \mu_\xi = 0, \sigma^2_\xi = 1, \sigma^2_\epsilon = 0.002, \sigma^2_\delta = 0.010$

Bias

	OLS		MALS	SLS
$n = 200$	theoretic	simulation	simulation	simulation
β_0	-0.005	-0.0044	-0.0000	0.0003
β_1	-0.010	-0.0100	-0.0007	0.0004
β_2	0.010	0.0095	-0.0000	-0.0002
$n = 900$	theoretic	simulation	simulation	simulation
β_0	-0.005	-0.0048	0.0001	0.0002
β_1	-0.010	-0.0100	-0.0001	-0.0000
β_2	0.010	0.0097	-0.0001	-0.0001

Standard Deviation

	OLS		MALS	SLS
$n = 200$	theoretic	simulation	simulation	simulation
β_0	0.0128	0.0123	0.0125	0.0093
β_1	0.0145	0.0143	0.0145	0.0121
β_2	0.0124	0.0118	0.0121	0.0092
$n = 900$	theoretic	simulation	simulation	simulation
β_0	0.0061	0.0059	0.0060	0.0045
β_1	0.0068	0.0065	0.0066	0.0056
β_2	0.0059	0.0054	0.0056	0.0042

Difference of Covariance Matrices

n=200

$$\frac{n}{\sigma_\delta^2}(\Sigma_{MALS} - \Sigma_{OLS}) \quad \begin{pmatrix} 0.093 & 0.026 & -0.094 \\ 0.026 & 0.081 & -0.036 \\ -0.094 & -0.036 & 0.128 \end{pmatrix}$$

$$\frac{n}{\sigma_\delta^2}(\Sigma_{MALS} - \Sigma_{SLS}) \quad \begin{pmatrix} 1.397 & 0.580 & -1.306 \\ 0.580 & 1.241 & -0.566 \\ -1.306 & -0.566 & 1.223 \end{pmatrix}$$

n=900

$$\frac{n}{\sigma_\delta^2}(\Sigma_{MALS} - \Sigma_{OLS}) \quad \begin{pmatrix} 0.068 & -0.008 & -0.068 \\ -0.008 & 0.175 & -0.079 \\ -0.068 & -0.079 & 0.140 \end{pmatrix}$$

$$\frac{n}{\sigma_\delta^2}(\Sigma_{MALS} - \Sigma_{SLS}) \quad \begin{pmatrix} 1.369 & 0.554 & -1.283 \\ 0.554 & 1.167 & -0.540 \\ -1.283 & -0.540 & 1.209 \end{pmatrix}$$

Cubic model: $\beta = (0, 1, -0.5, 0.5)', \mu_\xi = 0, \sigma_\xi^2 = 1, \sigma_\epsilon^2 = 0.002, \sigma_\delta^2 = 0.010$

Bias

	OLS		MALS	SLS
$n = 200$	theoretic	simulation	simulation	simulation
β_0	-0.005	-0.0042	-0.0002	0.0002
β_1	0.005	0.0016	-0.0086	-0.0021
β_2	0.010	0.0094	0.0002	0.0002
β_3	-0.015	-0.0133	0.0028	0.0010
$n = 900$	theoretic	simulation	simulation	simulation
β_0	-0.005	-0.0054	-0.0008	0.0002
β_1	0.005	0.0049	-0.0017	-0.0013
β_2	0.010	0.0103	0.0008	-0.0002
β_3	-0.015	-0.0146	0.0004	0.0005

Standard Deviation

	OLS		MALS	SLS
$n = 200$	theoretic	simulation	simulation	simulation
β_0	0.0396	0.0318	0.0324	0.0136
β_1	0.0961	0.0672	0.0693	0.0299
β_2	0.0522	0.0452	0.0463	0.0259
β_3	0.0406	0.0313	0.0324	0.0190
$n = 900$	theoretic	simulation	simulation	simulation
β_0	0.0187	0.0174	0.0178	0.0063
β_1	0.0453	0.0384	0.0398	0.0144
β_2	0.0246	0.0227	0.0235	0.0119
β_3	0.0191	0.0163	0.0169	0.0089

Difference of Covariance Matrices

n=200

$$\frac{n}{\sigma_\delta^2}(\Sigma_{MALS} - \Sigma_{OLS}) \qquad \begin{pmatrix} 0.795 & -0.493 & -1.163 & 0.157 \\ -0.493 & 5.720 & 0.609 & -2.620 \\ -1.163 & 0.609 & 1.955 & -0.252 \\ 0.157 & -2.620 & -0.252 & 1.481 \end{pmatrix}$$

$$\frac{n}{\sigma_\delta^2}(\Sigma_{MALS} - \Sigma_{SLS}) \qquad \begin{pmatrix} 17.368 & -11.793 & -22.198 & 5.765 \\ -11.793 & 78.149 & 12.380 & -32.179 \\ -22.198 & 12.380 & 29.421 & -6.421 \\ 5.765 & -32.179 & -6.421 & 13.765 \end{pmatrix}$$

n=900

$$\frac{n}{\sigma_\delta^2}(\Sigma_{MALS} - \Sigma_{OLS}) \qquad \begin{pmatrix} 1.322 & -0.288 & -2.041 & 0.157 \\ -0.288 & 9.246 & 0.361 & -3.650 \\ -2.041 & 0.361 & 3.349 & -0.340 \\ 0.157 & -3.650 & -0.340 & 1.744 \end{pmatrix}$$

$$\frac{n}{\sigma_\delta^2}(\Sigma_{MALS} - \Sigma_{SLS}) \qquad \begin{pmatrix} 25.017 & -18.048 & -30.203 & 7.438 \\ -18.048 & 123.641 & 18.954 & -47.745 \\ -30.203 & 18.954 & 36.927 & -7.876 \\ 7.438 & -47.745 & -7.876 & 18.647 \end{pmatrix}$$

8 Conclusion

We compared three estimators of the parameter vector β of a polynomial regression with measurement error: the naive ordinary least squares (OLS), the adjusted least squares (ALS), and the structural least squares (SLS) estimators. Although OLS is inconsistent it is still worthwhile to compare it with the consistent ALS and SLS estimators because often the standard errors of OLS are so small as compared to those of ALS and SLS that in certain occasions OLS might be preferable despite its inconsistency. SLS relies on the knowledge of the regressor distribution (structural case), which here is taken to be a normal distribution. ALS is not based on any distributional assumptions about the regressor (functional case) and is therefore more robust than SLS. Here, however, we do not study robustness properties (for this see Schneeweiss and Nittner (2000)). Instead we assume that SLS takes the normal regressor distribution correctly into account. In this case one might suppose that SLS is more efficient than ALS.

In order to study efficiency properties of the estimators we compute their asymptotic covariance matrices. As they are hard to compare in general we restrict our investigations to borderline cases with small error variances. Such cases seem to be quite realistic in practical applications. We study two scenarios: in the first, the measurement error variance σ_δ^2 tends to zero, in the second both error variances, σ_δ^2 and σ_ϵ^2, the variance of the error in the equation, tend to zero. For both cases approximate formulas of the asymptotic covariance matrices are derived which then can be compared to each other.

For small σ_δ^2 it turns out that, surprisingly, $\Sigma_{ALS} = \Sigma_{SLS}$ up to the order of σ_δ^2, whereas Σ_{OLS} differs clearly from Σ_{ALS} and Σ_{SLS}. We can, however, not say that $\Sigma_{OLS} < \Sigma_{ALS}$. In fact there are linear combinations $\alpha = a'\beta$ for which the variance of $\hat{\alpha}_{OLS}$ is larger than the variance of $\hat{\alpha}_{SLS}$. On the other hand, for the components of $\Phi\beta$, where $\Phi = E\zeta\zeta'$, we can say that the OLS estimator has smaller variance than the ALS (or SLS) estimator if σ_δ^2 is small enough.

In the case where σ_ϵ^2 and σ_δ^2 are both small it turns out, again surprisingly, that $\Sigma_{OLS} = \Sigma_{ALS}$ up to the order of σ_δ^2, whereas now Σ_{SLS} differs from Σ_{ALS} and Σ_{OLS}. In fact we have essentially that $det\Sigma_{ALS} > det\Sigma_{SLS}$ in the limit, see Kukush and Schneeweiss (2000).

We not only have these qualitative results, we also have explicit quantitative formulas for the covariance matrices (see (15), (25), (38), (41)) and for the bias of OLS (see (14)) up to the order of σ_δ^2. These formulas can be used to compute approximately these covariance matrices. Simulations show that these approximations are fairly accurate for small error variances. It might be advantageous, though, to replace the matrix $\Phi = E\zeta\zeta'$ with Ezz' in these computations, but we did not try out this modification.

Appendix: Convergence of iteratively reweighted least squares

We investigate the convergence properties of "iteratively reweighted least squares", which is an iterative procedure intended to solve the estimating equations for $\hat{\beta}_{SLS}$ and $\hat{\sigma}_{\epsilon SLS}$. For convenience we repeat the estimating equations of Section 4.1 in a slightly different form:

$$\frac{1}{n}\sum_1^n \frac{y_i - \hat{\mu}(x_i)'\beta}{\hat{v}(x_i,\beta,\sigma_\epsilon^2)}\ \hat{\mu}(x_i) = 0 \tag{42}$$

$$\sigma_\epsilon^2 = \frac{1}{n-k-1}\sum_1^n \{y_i - \hat{\mu}(x_i)'\beta\}^2 - \frac{1}{n}\sum_1^n \beta' A(x_i)\beta\ , \tag{43}$$

where $A(x_i) = \hat{M}(x_i) - \hat{\mu}(x_i)\hat{\mu}(x_i)'$.

Let Θ_β be a closed bounded subset of $I\!R^{k+1}$ and $\Theta = \Theta \times [a,b]$ with $0 < a < b < \infty$. We suppose that the true values β^0 and $\sigma_{\epsilon 0}^2$ satisfy the following condition: β^0 is an interior point of Θ_β, and

$$\sigma_{\epsilon 0}^2 \in (a,b) \tag{44}$$

In Kukush and Schneeweiss (2000) it is shown that under this condition the system (42), (43) has a solution $\hat{\beta}_{SLS}$, $\hat{\sigma}_{\epsilon SLS}^2$ a.s. for all $n \geq n_0(\omega)$, and $\hat{\beta}_{SLS} \to \beta^0$, $\hat{\sigma}_{\epsilon SLS}^2 \to \sigma_{\epsilon 0}$ a.s., as $n \to \infty$. (Actually this was shown for a slightly different system of estimating equations, but the arguments are still valid for (42), (43)). However, this was only an existence not a uniqueness proof and no procedure to actually find a solution was provided.

Here we consider the following iterative procedure to solve (42), (43) uniquely. From (42) derive the partial solution

$$\begin{aligned} \beta &= \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\hat{\mu}(x_i)\hat{\mu}'(x_i)}{\hat{v}(x_i, \beta, \sigma_\epsilon^2)} \right)^{-1} \cdot \frac{1}{n} \sum_{i=1}^{n} \frac{y_i \hat{\mu}(x_i)}{\hat{v}(x_i, \beta, \sigma_\epsilon^2)} \\ &\stackrel{df}{=} \varphi(\beta, \sigma_\epsilon^2) \end{aligned} \tag{45}$$

Note that as $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\hat{\mu}(x_i)\hat{\mu}'(x_i)}{\hat{v}(x_i, \beta, \sigma_\epsilon^2)} \to E_0 \frac{\mu(x)\mu'(x)}{v(x, \beta, \sigma_\epsilon^2)} \tag{46}$$

uniformly on Θ a.s. , see Kukush and Schneeweiss (2000), proof of Lemma 1. The limit matrix in (46) is positive definite. Therefore the function $\varphi : \Theta \to I\!R^{k+1}$ in (45) is well defined a.s. for $n \geq n_1(\omega)$.

Denote also the function on the right hand side of (43) by $\psi(\beta)$, where $\psi : \Theta_\beta \to I\!R$. Then the system of equations (45), (43) can be written as

$$\beta = \varphi(\beta, \sigma_\epsilon^2) \ , \tag{47}$$

$$\sigma_\epsilon^2 = \psi(\beta) \ . \tag{48}$$

Introduce the projector P on the interval $[a, b]$,

$$P(z) = \begin{cases} z & , \text{ if } \ z \in [a, b] \\ a & , \text{ if } \ z \in (-\infty, a) \\ b & , \text{ if } \ z \in (b, +\infty) \ . \end{cases}$$

We modify (48) to

$$\sigma_\epsilon^2 = P \circ \psi(\beta) \ . \tag{49}$$

Now the following iterative algorithm is proposed:

1. Given estimates $\beta^{(j)} \in \Theta_\beta$ and $\sigma_\epsilon^{2(j)} \in [a, b]$ from the j-th iteration of the algorithm, find

$$\sigma_\epsilon^{2(j+1)} = P \circ \psi(\beta^{(j)}) \ . \tag{50}$$

2. Find

$$\beta^{(j+1)} = \varphi(\beta^{(j)}, \sigma_\epsilon^{2(j+1)}). \tag{51}$$

The initial value $\beta^{(0)} \in \Theta_\beta$ can be chosen arbitrarly (e.g., as the naive OLS estimate), and we get $\sigma_\epsilon^{2(0)} = P \circ \psi(\beta^{(0)})$. This algorithm leads to the desired solution.

Lemma: The system (47), (48) has a unique solution $\hat{\beta}_{SLS}$, $\hat{\sigma}^2_{\epsilon SLS}$ a.s. for $n \geq n_2(w)$ and

$$\lim_{n\to\infty} \beta^{(j)} = \hat{\beta}_{SLS} \ , \ \lim_{n\to\infty} \sigma_\epsilon^{2(j)} = \hat{\sigma}^2_{\epsilon SLS} \text{ a.s.}$$

Proof: The proof consists of several steps.
1. Substituting (54) into (47) we define a new function $F : \Theta_\beta \to I\!R^{k+1}$ by

$$F(\beta) = \varphi(\beta, P \circ \psi(\beta)) \ . \tag{52}$$

The system of equations (47), (54) is equivalent to the system

$$\beta = F(\beta) \tag{53}$$

$$\sigma_\epsilon^2 = P \circ \psi(\beta) \tag{54}$$

We want to show that the function F thus defined is a contraction mapping acting on the compact set Θ_β.
2. We show: The partial derivatives of φ tend to zero. Denote

$$A_n = \frac{1}{n}\sum_1^n \frac{\hat{\mu}(x_i)\hat{\mu}'(x_i)}{\hat{v}(x_i,\beta,\sigma_\epsilon^2)} \quad , \quad B_n = \frac{1}{n}\sum_1^n \frac{\hat{\mu}(x_i)\zeta_i'}{\hat{v}(x_i,\beta,\sigma_\epsilon^2)} \ .$$

Then from (45) we have

$$\varphi(\beta,\sigma_\epsilon^2) = A_n^{-1}B_n\beta^0 + A_n^{-1}\frac{1}{n}\sum_1^n \frac{\epsilon_i\hat{\mu}(x_i)}{\hat{v}(x_i,\beta,\sigma_\epsilon^2)} \ .$$

Denote $\theta = (\beta', \sigma_\epsilon^2)'$. For every partial derivative $\frac{\partial}{\partial\theta_j}$ we have

$$\begin{aligned}\frac{\partial\varphi(\theta)}{\partial\theta_j} &= \left(-A_n^{-1}\frac{\partial A_n}{\partial\theta_j}A_n^{-1}B_n + A_n^{-1}\frac{\partial B_n}{\partial\theta_j}\right)\beta^0 \\ &- A_n^{-1}\frac{\partial A_n}{\partial\theta_j}A_n^{-1}\cdot\frac{1}{n}\sum_1^n\frac{\epsilon_i\hat{\mu}(x_i)}{\hat{v}(x_i,\theta)} \\ &+ A_n^{-1}\frac{1}{n}\sum_1^n \epsilon_i\hat{\mu}(x_i)\cdot\left(-\frac{\frac{\partial\hat{v}(x_i,\theta)}{\partial\theta_j}}{\hat{v}^2(x_i,\theta)}\right) \ .\end{aligned} \tag{55}$$

Now, uniformly on Θ a.s. we have as $n \to \infty$

$$
\begin{aligned}
A_n &\to E_0 \tfrac{\mu(x)\mu'(x)}{v(x,\theta)} \ , \\
B_n &\to E_0 \tfrac{\mu(x)\mu'(x)}{v(x,\theta)} \ , \\
\tfrac{\partial A_n}{\partial \theta_j} &= \tfrac{1}{n}\sum_1^n \hat{\mu}(x_i)\hat{\mu}'(x_i)\tfrac{\partial}{\partial \theta_j}\left(\hat{v}^{-1}(x_i,\theta)\right) \to E_0 \mu(x)\mu'(x)\tfrac{\partial}{\partial \theta_j}\left(v^{-1}(x,\theta)\right) , \\
\tfrac{\partial B_n}{\partial \theta_j} &= \tfrac{1}{n}\sum_1^n \hat{\mu}(x_i)\zeta_i'\tfrac{\partial}{\partial \theta_j}\left(\hat{v}^{-1}(x_i,\theta)\right) \to E_0 \mu(x)\mu'(x)\tfrac{\partial}{\partial \theta_j}\left(v^{-1}(x,\theta)\right) .
\end{aligned}
$$

The limits of B_n and of $\partial B_n/\partial \theta_j$ follow from the fact that $E(\zeta|x) = \mu(x)$. All four limits imply that uniformly on Θ a.s.

$$
-A_n^{-1}\frac{\partial A_n}{\partial \theta_j}A_n^{-1}B_n + A_n^{-1}\frac{\partial B_n}{\partial \theta_j} \to 0 \ .
$$

The other terms in (55) also converge to 0 uniformly on Θ a.s. Therefore

$$
\lim_{n\to\infty}\sup_{\theta\in\Theta}\left\|\frac{\partial \varphi(\theta)}{\partial \theta'}\right\| = 0 \ , \ a.s. \tag{56}
$$

3. The partial derivatives of ψ are bounded. Indeed, as Θ_β is a compact set we have

$$
\sup_{n\geq k+2}\sup_{\beta\in\Theta_\beta}\left\|\frac{\partial \psi(\beta)}{\partial \beta}\right\| < \infty \ a.s. \tag{57}
$$

4. The mapping F is a contraction. First note that without loss of generality we may assume that Θ_β is convex - if not, simply take the convex hull of Θ_β as the new Θ_β. For $\beta_1, \beta_2 \in \Theta_\beta$ we then have:

$$
\|F(\beta_1) - F(\beta_2)\| = \left\|\varphi(\beta_1, \sigma_{\epsilon 1}^2) - \varphi(\beta_2, \sigma_{\epsilon 2}^2)\right\| \ .
$$

Here $\sigma_{\epsilon i}^2 \overset{df}{=} P \circ \psi(\beta_i)$, $i = 1, 2$. Next,

$$
\|F(\beta_1) - F(\beta_2)\| \ \leq \ \sup_{\theta\in\Theta}\left\|\frac{\partial \varphi(\theta)}{\partial \theta'}\right\| \sqrt{\|\beta_1 - \beta_2\|^2 + (\sigma_{\epsilon 1}^2 - \sigma_{\epsilon 2}^2)^2}
$$

We have for $n \geq k+2$

$$
\begin{aligned}
\left|\sigma_{\epsilon 1}^2 - \sigma_{\epsilon 2}^2\right| &= |P \circ \psi(\beta_1) - P \circ \psi(\beta_2)| \leq |\psi(\beta_1) - \psi(\beta_2)| \\
&\leq \sup_{n\geq k+2}\sup_{\beta\in\Theta_\beta}\left\|\frac{\partial \psi(\beta)}{\partial \beta}\right\| \|\beta_1 - \beta_2\| \ .
\end{aligned}
$$

Therefore

$$\begin{aligned} \|F(\beta_1) - F(\beta_2)\| &\leq \sup_{\theta \in \Theta} \left\| \frac{\partial \varphi(\theta)}{\partial \theta'} \right\| \left(1 + \sup_{n \geq k+2} \sup_{\beta \in \Theta_\beta} \left\| \frac{\partial \psi(\beta)}{\partial \beta} \right\| \right) \|\beta_1 - \beta_2\| \\ &= \lambda_n \|\beta_1 - \beta_2\| \ . \end{aligned} \tag{58}$$

But according to (56) and (57), $\lambda_n \to 0$ as n $\to \infty$ a.s.

5. The mapping F is a contraction $\Theta_\beta \to \Theta_\beta$. To prove this we use the fact that $\hat{\theta} = (\hat{\beta}'_{SLS}, \hat{\sigma}^2_{\epsilon SLS})'$ is a consistent solution of (47), (48), see Kukush and Schneeweiss (2000). As $\hat{\theta}$ converges a.s. to an interior point of Θ, we have that $\hat{\theta}$ will be also an interior point of Θ a.s. for $n \geq N(\omega)$, and for some $r = r(\omega) > 0$, which does not depend on n, we will have

$$B(\hat{\beta}_{SLS}, r(\omega)) \subset \Theta_\beta \quad \text{for all } n \geq N(\omega) \ ,$$

where $B(\beta, r)$ denotes a ball with center β and radius r. Then, for $n \geq N(\omega)$, $\hat{\theta}$ will also be a solution of (47), (54), and $\hat{\beta}_{SLS}$ will be a solution of (53):

$$F(\hat{\beta}_{SLS}) = \hat{\beta}_{SLS} \ . \tag{59}$$

For each $\beta \in \Theta_\beta$ we now have, due to (58) and (59),

$$\left\| F(\beta) - \hat{\beta}_{SLS} \right\| \leq \lambda_n \left\| \beta - \hat{\beta}_{SLS} \right\| \leq \lambda_n \max_{\beta_1, \beta_2 \in \Theta_\beta} \|\beta_1 - \beta_2\| \ .$$

But $\lambda_n \to 0$ a.s., therefore for $n \geq N_1(\omega)$

$$F(\beta) \in B(\hat{\beta}_{SLS}, r(\omega)) \subset \Theta_\beta \ .$$

Thus for $n \geq N_2(\omega)$, $F : \Theta_\beta \to \Theta_\beta$, and in addition, due to (58),

$$\|F(\beta_1) - F(\beta_2)\| \leq \frac{1}{2} \|\beta_1 - \beta_2\| \ .$$

Hence F is a contraction on Θ_β for sufficiently large n .

6. Because of (59) $\hat{\beta}_{SLS}$ is a fixed point of the contraction F. Applying Banach's Theorem, it now follows that a.s. for $n \geq N_2(\omega)$ $\hat{\beta}_{SLS}$ is the unique solution of the equation $\beta = F(\beta)$, $\beta \in \Theta_\beta$, and any sequence $\beta^{(j+1)} = F(\beta^{(j)})$, $\beta^{(0)} \in \Theta_\beta$, converges to the solution $\hat{\beta}_{SLS}$.

7. Having found a convergent sequence $\beta^{(j)} \to \hat{\beta}_{SLS}$, we can substitute these $\beta^{(j)}$ into (50) and find that for all $n \geq N_2(\omega)$

$$\sigma_\epsilon^{2(j+1)} \to P \circ \psi(\hat{\beta}_{SLS}) = \hat{\sigma}^2_{\epsilon SLS} \; .$$

This proves the lemma.

Remark. Suppose that $\lambda_n < 1$, and $\lambda_n \cdot \max_{\beta_1, \beta_2 \in \Theta_\beta} \|\beta_1 - \beta_2\| \leq r(\omega)$. Then

$$\begin{aligned} \left\|\beta^{(j)} - \hat{\beta}_{SLS}\right\| &\leq \sum_{p=j}^{\infty} \left\|F(\beta^{(j)}) - F(\beta^{(j+1)})\right\| \leq \sum_{p=j}^{\infty} \lambda_n^p \cdot \left\|\beta^{(0)} - \beta^{(1)}\right\| \\ &= \frac{\lambda_n^j}{1-\lambda_n} \cdot \left\|\beta^{(0)} - \beta^{(1)}\right\| \; . \end{aligned}$$

References

CARROLL, R. J., RUPPERT, D. and STEFANSKI, L. A. (1995). *Measurement Error in Nonlinear Models.* Chapman and Hall, London.

CHENG, C.-L. and SCHNEEWEISS, H. (1998). *Polynomial regression with errors in the variables.* J. R. Statist. Soc. B **60**, 189–199.

CHENG, C.-L., SCHNEEWEISS, H. and THAMERUS, M. (2000). *A small sample estimator for a polynomial regression with errors in the variables.* J. R. Statist. Soc. B **62**, 699–709.

KUHA, J. T. and TEMPLE, J. (1999). *Covariate measurement error in quadratic regression.* Discussion Paper 1999-W2, Nuffield College Oxford.

KUKUSH, A. and SCHNEEWEISS, H. (2000). *A Comparison of Asymptotic Covariance Matrices of Adjusted Least Squares and Structural Least Squares in Error Ridden Polynomial Regression.* Discussion Paper 218, Sonderforschungsbereich 386, University of Munich.

SCHNEEWEISS, H. and NITTNER, T. (2000). *Estimating a polynomial regression with measurement errors in the structural and in the functional case - a comparison.* Discussion Paper 197, Sonderforschungsbereich 386, University of Munich.

STEFANSKI, L. A. (1989). *Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models.* Communications in Statistics, Part A - Theory and Methods, **18**, 4335-4358.

THAMERUS, M. (1998). *Different nonlinear regression models with incorrectly observed covariates.* In R. Galata and H. Kuechenhoff (Eds.): Econometrics in Theory and Practice, Festschrift for Hans Schneeweiss. Physica, Heidelberg, New York.